Acceptor substrate discrimination in phosphatidyl-myo-inositol mannoside synthesis: Structural and mutational analysis of mannosyltransferase Corynebacterium glutamicum PimB'

Batt, Sarah; Jabeen, Talat; Mishra, Arun Kumar; Veerapen, Natacha; Krumbach, K; Eggeling, L; Besra, Gurdyal; Futterer, Klaus

DOI:
10.1074/jbc.M110.165407

JBC Papers in Press. Published on September 15, 2010 as Manuscript M110.165407
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M110.165407

ACCEPTOR SUBSTRATE DISCRIMINATION IN PHOSPHATIDYL-*MYO*-INOSITOL MANNOSIDE SYNTHESIS: STRUCTURAL AND MUTATIONAL ANALYSIS OF MANNOSYLTRANSFERASE *CORYNEBACTERIUM GLUTAMICUM* PIMB'

Sarah M. Batt[1], Talat Jabeen[1], Arun K. Mishra[1], Natacha Veerapen[1], Karin Krumbach[2], Lothar Eggeling[2], Gurdyal S. Besra[1]*, Klaus Fütterer[1]*

From the School of Biosciences[1], University of Birmingham, Edgbaston, Birmingham B15 2TT, UK, and the Institut für Biotechnologie I[2], Forschungszentrum Jülich, D-52425 Jülich, Germany.

Running Head: Structural and mutational analysis of PimB'

*Address correspondence to Klaus Fütterer or Gurdyal S. Besra, School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK. Phone +44 –121–4145895; Fax +44–121–4145925; email K.Futterer@bham.ac.uk or G.Besra@bham.ac.uk.

Long-term survival of the pathogen *Mycobacterium tuberculosis* in humans is linked to the immunomodulatory potential of its complex cell wall glycolipids, which include the phosphatidylinositol mannoside (PIM) series as well as the related lipomannan and lipoarabinomannan glycoconjugates. PIM biosynthesis is initiated by a set of cytosolic α-mannosyltransferases, catalysing glycosyl transfer from the activated saccharide donor GDP-Man to the acceptor phosphatidyl-*myo*-inositol (PI) in an ordered and regio-specific fashion. Herein, we report the crystal structure of mannosyltransferase *Corynebacterium glutamicum* PimB' in complex with nucleotide to a resolution of 2.0 Å. PimB' attaches mannosyl selectively to the 6-OH of the inositol moiety of PI. Two crystal forms and GDP- versus GDP-Man-bound complexes reveal flexibility of the nucleotide conformation as well as of the structural framework of the active site. Structural comparison, docking of the saccharide acceptor and site directed mutagenesis pin regio-selectivity to a conserved Asp residue in the N-terminal domain that forces presentation of the correct inositol hydroxyl to the saccharide donor.

The cell envelope of *Mycobacterium tuberculosis*, the infectious agent causing tuberculosis (TB), contains a variety of glycolipids that play a central role in subverting the host's immune response, and thus help establish a long-lasting latent infection, a hallmark of the pathophysiology of TB (1, 2). Phosphatidylinositol mannosides (PIM) represent a series of glycolipids that comprise a phosphatidyl-*myo*-inositol (PI) core, an acylated mannosyl group attached to the 2-hydroxyl of inositol, and a mannosyl-oligosaccharide of variable length attached to the inositol 6-hydroxyl (Fig. 1) (3). A precursor of the more complex lipomannan (LM) and lipoarabinomannan (LAM) glycolipids, PIM have been shown to influence both innate (4, 5) and adaptive immunity of the host (6-9).

PIM biosynthesis begins by consecutive transfer of two mannosyl-units from activated sugar-nucleotide (GDP-Man) to PI, catalyzed by cytoplasmic α-mannosyltransferases (Fig. 1). Attachment of the first mannosyl residue to the 2-hydroxyl of the inositol ring, resulting in PIM_1, is catalyzed by PimA (*M. tuberculosis* Rv2610c, *Mycobacterium smegmatis* MSMEG_2935) (10, 11), followed by acylation of the 2-mannose (by *M. tuberculosis* Rv2611c) (12) to yield mono-acylated PIM_1 (Ac_1PIM_1). The second mannosyl residue is attached at the 6-hydroxyl, to yield Ac_1PIM_2, a reaction catalyzed by PimB' (*M. tuberculosis* Rv2188c, *M. smegmatis* MSMEG_4253, *Corynebacterium glutamicum* NCgl2106) (13-15). The designation PimB had originally been assigned to open reading frame Rv0557, also encoding an α-mannosyltransferase. Subsequently, Rv0557 was found to be dispensable for synthesis of Ac_1PIM_2 and the corresponding protein has since been renamed MgtA (15-17). For consistency with the recent literature, we retain the designation PimB' for Rv2188c (and its orthologs in *M. smegmatis* and *C. glutamicum*) (14, 15, 18).

Prolonged incubation of Ac_1PIM_2 with PimB' or PimA does not lead to further

extension of Ac_1PIM_2 (14). Instead, extending the mannosyl chain at the 6-OH requires a distinct set of mannosyltransferases. For instance, bioinformatical analysis of the genome of *M. tuberculosis* CDC1551 led to the identification of PimC, which catalyses synthesis of the trimannoside Ac_3PIM_3 (19). Still, the *pimC* deletion in *Mycobacterium bovis* bacille Calmette-Guérin (BCG) does not interrupt formation of higher PIMs, LM or LAM, for which $PIM_{n=1-3}$ is considered a precursor, while genes orthologous to *pimC* were found in only 22% of clinical isolates. Thus, compensatory synthetic pathways must exist. Higher order PIM, LM, and LAM depend on glycosyl transfer from the lipid-saccharide donor C_{30}/C_{50}-P-Man and membrane-embedded glycosyltransferases, including PimE (20), with evidence for pathway bifurcation at Ac_2PIM_4 (21, 22).

The three-dimensional structures of soluble glycosyltransferases display only two fundamental fold topologies, termed GT-A and GT-B (23), contrasting with the diversity of protein folds among glycoside hydrolase enzymes (24). According to the sequence-based classification of carbohydrate-active enzymes, PimB' (Rv2188c) belongs to glycosyltransferase family 4 (GT4, http://www.cazy.org/GT4.html (25)). Encompassing a diverse range of enzymatic activities and substrate specificities, GT4 family enzymes assume the GT-B fold, with known structures for about half a dozen family members (26-30). The GT-B topology is characterized by two Rossmann fold-like domains, where donor and acceptor substrates bind in the central cleft between the two domains. In GT4 family transferases, the C-terminal domain provides the majority of contacts for the nucleoside-diphosphate-saccharide donor substrate, while the diverse acceptor-substrates bind to the more variable N-terminal domain. A hinge region allows for conformational flexibility between the two domains, which can be dramatic, as is illustrated by the structures of ligand-free and substrate-bound *C. glutamicum* MshA (PDB entries 3c48, 3c4v (29)). Between these two states, the relative orientation of the domains changes by a 97º rotational movement. Similar if less dramatic examples of inter-domain flexibility have been observed in structures of several other GT4 family members.

Recently determined structures of GT4 family enzymes includes that of *M. smegmatis* PimA, the enzyme catalysing transfer of the first mannosyl group to PI (27). This structure shed light on the mode of nucleotide binding and suggested a model for recognition of the acceptor substrate. While PimA and PimB' utilize the same donor substrate, their acceptor specificity is distinct. PimB' appears unable to mannosylate phosphatidyl-*myo*-inositol (or *myo*-inositol-1-phosphate) on the 2-hydroxyl, while PimA does not mannosylate PI or PIM_1 on the 6-hydroxyl.

In order to clarify structural differences between PimA and PimB' that could explain regio-selectivity we undertook the structure determination of PimB'. Attempts to generate recombinant protein of *M. tuberculosis* PimB' (Rv2188c) were unsuccessful, but overexpression of the *C. glutamicum* ortholog (NCgl2106) in *Escherichia coli* yielded soluble protein that crystallized when incubated with the donor substrate. Herein, we report the crystal structure of *C. glutamicum* PimB' in complex with nucleotide to a resolution of 2.0 Å, revealing flexibility of both the nucleotide conformation and the structural framework of the active site. The results of our site-directed mutational analysis are consistent with a substrate-mediated S_Ni (internal return) reaction mechanism, and indicate that a conserved aspartic acid in the acceptor-binding domain is critical in determining regio-selectivity of mannosyl transfer.

EXPERIMENTAL PROCEDURES

Recombinant protein production– Cloning of *C. glutamicum* PimB' has been described previously (15). Liquid cultures of E. coli BL21 (DE3), harboring pET16b-*pimB'*, were grown at 30ºC in LB-broth medium (Difco), supplemented with ampicillin (100 μg/mL). Expression of *pimB'* was induced at OD_{600} 0.4 – 0.6 by adding 0.5 mM isopropyl-β-D-thiogalactoside (IPTG). Cultures were incubated for a further 4 hours, before being harvested (4ºC, 4000*g*), washed with 0.85% saline and stored at –20 ºC. For the expression of SeMet-labeled PimB', liquid cultures were grown to OD_{600} 0.4, then pelleted and resuspended in 1L minimal media twice (50 mM Na_2HPO_4, 20 mM KH_2PO_4, 10 mM NaCl, 20 mM NH_4Cl, 20 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 0.4% (w/v) glucose). Cells were grown at 37ºC to OD_{600} 0.5,

then adding 100 mg L-lysine, L-phenylalanine and L-threonine, and 50 mg of isoleucine, leucine and valine, and 60 mg of L-seleno-methionine. After 30 minutes at 37ºC, expression of *pimB'* was induced adding 0.5 mM IPTG, followed by incubation at 25ºC for 24 hours. Cells were harvested and treated as before.

Purification and crystallization– Cell pellets were thawed on ice and resuspended in 30 mL of 50 mM NaH_2PO_4, pH 8, 300 mM NaCl and 10 mM imidazole (Buffer A), supplemented with an EDTA-free protease cocktail tablet (Roche). The suspension was sonicated (Sonicator Ultrasonic Liquid Processor XL, Misonix Inc.) on ice for a total time of 10 minutes, using 20-second pulses and 40-second cooling periods. The lysate was centrifuged (27000*g*, 30 minutes, 4°C) and the supernatant passed through a Buffer A-equilibrated Ni^{2+}-charged His-Trap column (1 mL, GE Healthcare). The column was subsequently washed with 50 mL of Buffer A with 20 mM imidazole, and the protein eluted with a 50-300 mM step gradient of imidazole. Fractions containing protein were dialyzed against 20 mM Tris-HCl pH 7.5, 10 mM NaCl, 10% glycerol and 1 mM DTT. After dialysis, proteins were concentrated by ultrafiltration to 8 mg/mL using a 10 kDa cut-off (Centriprep, Millipore).

Prior to setting up crystallization trials, N-terminally tagged PimB' was incubated overnight with 10 mM GDP-Man. Commercial sparse matrix screens (Molecular Dimensions) were used to identify crystallization conditions in hanging drop vapor diffusion experiments. Thin plate crystals formed at 6-26% PEG 3350, 0.1 M Bis-Tris pH 5.5, 0.1 M lithium sulfate and 0.1 M glycine. Crystals were cryoprotected by adding reservoir solution supplemented with 8% w/v sucrose and 8% v/v glycerol to the crystallization drop with subsequent flash freezing in liquid nitrogen. Crystals of N-terminally tagged PimB' were in space group *P*1, diffracting to 3.5 Å on the in-house X-ray source (Rigaku MicroMax 007HF, Saturn 944 CCD), and to 2.2 Å at the synchrotron.

Structure determination of triclinic crystal form– Single wavelength data of native *C. glutamicum* PimB' were recorded on beamline ID29 at ESRF, Grenoble, France (Table 1) at 10% transmission. Analysis of the Matthews volume suggested the presence of two molecules of PimB' per asymmetric unit (V_M = 2.28 Da/Å^3) with a predicted solvent fraction of 47%. Initial attempts to phase by molecular replacement (MR) resulted in a statistically significant solution in terms of the PHASER Z-score (*Z* = 9.0, resolution range 10 - 3Å) (31), but failed to yield a useful electron density map. As the thin plate crystals did not tolerate soaking in heavy metal salts, we generated SeMet-derivatized crystals of PimB' and recorded data at the Se-K edge (Table 1). Significant radiation damage of the cryogenically cooled crystals limited data acquisition to the peak wavelength, and due to the triclinic symmetry, resulted in data of low multiplicity (Table 1). In order to optimize the anomalous signal and completeness of Bijvoet pairs an inverse beam protocol was used, recording data in sets of 20 successive frames (0.5º oscillation) at ϕ and ϕ + 180º, respectively. Data were reduced using XDS and XSCALE (32). Only one of several SeMet SAD data sets (SeMet1 in Table 1) recorded from a series of crystals provided a sufficiently strong anomalous signal to extract the Se positions using SHELXD (33). Patterson self-rotation maps indicated a non-crystallographic (NCS) 2-fold rotation axis parallel to the *b*-axis, consistent with the assumption of two molecules per asymmetric unit. Therefore, Se-positions obtained in runs of SHELXD were analyzed with PROFESSS (34) to derive the NCS operator and to verify that this operator was consistent with the self-rotation function. An initial set of 10 Se positions (5 per molecule) could be identified this way. SAD phasing in SHARP (35) allowed us to establish an additional set of 8 positions through difference Fourier analysis. Thus, 18 of 20 possible Se positions were obtained, but were not sufficient to obtain an interpretable map, either in SAD phasing using the best SeMet data or in phasing by MIR using the best SeMet together with the native data.

Next, we co-crystallized PimB' (N-terminally tagged) with a brominated derivative of GDP-Man, the synthesis of which has been described in (36). Combining the PimB' native data with this brominated derivative and the best four SeMet data sets led to a density map with recognizable secondary structure features (supplemental Fig. 2S*A*), albeit still too poor to build. Two-fold NCS averaging and phase extension to nominally 2.2 Å distinctly

improved the map (Supplemental Fig. 2S*B*). In order to support map interpretation we reconciled the previously established MR solution with the experimentally determined set of Se positions. This was aided by generating a homology model of the C-terminal domain of PimB' (using MODELLER (37)), which reproduced a subset of distance vectors between Se positions. Superimposing the homology model onto the experimental map aided interpretation of poorly defined areas, greatly accelerating model building. Iterative rounds of model building and minimization in REFMAC5 (38) very quickly led to a high quality 2Fo – Fc map (supplemental Fig. 2S*C*) that revealed previously undefined parts of the structure. The final model describes residues 4 to 381 of the PimB' (381 residues), with excellent stereochemistry (Table 1).

In both molecules of the triclinic crystals, the active site contains weak density for GDP-Man. Therefore, the ligand was omitted from the model until the structure refinement had converged to R_{cryst}/R_{free} values of 19.1%/22.4%. Adding GDP to this late-stage model, at full occupancy, reduced Rfree marginally, giving rise to negative difference density. Systematic variation of the occupancy of GDP between 0.5 and 1 resulted in a flat difference map around the ligand for 0.75. At this occupancy, the B-factors of the ligand atoms refined to an average of 31 $Å^2$, compared to a protein average of 12.5 $Å^2$ and an overall Wilson B-factor of 19.3 $Å^2$. While density for the mannose moiety is visible (Fig. 4*A*), it is insufficient to fit the saccharide unequivocally.

Structure determination of the orthorhombic crystal form– C-terminally tagged PimB' (8 mg/ml) was incubated overnight with either GDP or GDP-Man (20 mM) prior to setting up vapor diffusion experiments. Crystals of sufficient size appeared over reservoirs containing 0.1 M DL-malic acid pH 6.1, 13% PEG 3350 (GDP-Man complex) or 0.1 M sodium citrate pH 5.1, 22% PEG 3350 (GDP complex). Crystals were cryoprotected in mother liquor supplemented with 8% w/v sucrose and 8% glycerol and frozen in liquid nitrogen. The structure of the orthorhombic crystal form of PimB' was solved by MR, using a monomer of the refined 'triclinic' structure as search model in PHASER (31). The structural models were rebuilt in COOT and refined using REFMAC5 (38). Figures 2 – 5 were prepared using *PyMOL* (www.pymol.org).

Preparation of polar lipids and Ac_1PIM_1– Polar lipids containing Ac_1PIM_1 were extracted from *C. glutamicum-ΔpimB'ΔmgtA* as described previously (15, 39). The lipid extract was examined by 2-dimensional thin layer chromatography (TLC) on aluminum-backed plates of silica gel 60 F254 (Merck 5554), using $CHCl_3/CH_3OH/H_2O$ (60:30:6, v/v/v) in the first direction and $CHCl_3/CH_3COOH/CH_3OH/H_2O$ (40:25:3:6, v/v/v/v) in the second direction. The glycolipids were visualized by spraying plates with either α-naphthol/sulfuric acid, Dittmer & Lester Reagent, or 5% ethanolic molybdophosphoric acid followed by gentle charring of plates. Glycolipids were further separated into individual PIMs and other glycolipid species by preparative TLC on 10 cm × 20 cm plastic backed TLC plates of silica gel 60 F254 (Merck 5554), run in chloroform/methanol/water (60:30:6, v/v/v). The plates were then sprayed with 0.01% 1,6-diphenyl-1,3,5-hexatriene dissolved in petroleum ether/acetone (9:1 v/v) and the glycolipids were visualized under UV light. Following detection the plates were re-developed in toluene to remove diphenyl-1,3,5-hexatriene and the corresponding glycolipid bands were scraped from the plates and extracted from the silica gel using chloroform/methanol (2:1, v/v). Samples were analyzed by MALDI-TOF-MS as described previously (17).

Activity assay– PimB' was assayed for mannosyltransferase activity using a method adapted from (40). 20 µg of PimB' was incubated with either purified Ac_1PIM_1 (4 - 200 µg) or polar lipids (100 µg) extracted from *C. glutamicum-ΔpimB'ΔmgtA*, in the presence of 0.5 µCi GDP-[^{14}C]mannose (GE Healthcare), 0.125 mM ATP, 1 mM NADP, 10 mM $CaCl_2$, 0.1% (v/v) IGEPAL (Sigma) and 20 mM Tris-HCl pH 7.5 in a total volume of 100 µL. The reaction was incubated at 37°C for 30 minutes and quenched with the addition of $CHCl_3$:MeOH:H_2O (10:10:3). Unincorporated GDP-[^{14}C]mannose was removed by the addition of 2.625 mL of $CHCl_3$ and 1.125 mL of H_2O, followed by centrifugation to separate the organic and aqueous layers. The lower organic layer was washed three times with 3 mL of $CHCl_3$:MeOH:H_2O (3:47:48) and dried.

Incorporation of GDP-[^{14}C]mannose in the acceptor was measured using a scintillation counter and samples were analyzed by TLC run in $CHCl_3$:MeOH:H_2O:NH_4OH (65:25:3.6:0.5).

Trp fluorescence binding assay– Binding of GDP and GDP-Man was measured using intrinsic tryptophan fluorescence. The sample was excited at a wavelength of 295 nm and the resulting fluorescence emission was measured between 300 and 400 nm (scan speed 100 nm/min) using a Photon Technology International fluorescence spectrometer. Spectra were recorded for each addition of GDP-Man or GDP, to 700 μL of a 2 μM solution of PimB'. The experiment was repeated in triplicate and fluorescence intensity at 330 nm was plotted against ligand concentration and fitted to a one-site saturation binding model using SIGMAPLOT (Systat). In order to compensate for inner filtering of the ligand, we monitored fluorescence of albumin at a range of GDP concentrations, and derived a concentration-dependent correction to the Trp-fluorescence signal.

Docking of PIM_1– A molecular model of PIM_1 was manually docked into the active site of PimB' (closed configuration) guided by the following considerations: i) no steric clashes must occur and no adjustments to the protein structure should be necessary; ii) the 6-hydroxyl of inositol should be presented to C1 of the donor saccharide, while minimizing the oxygen-carbon distance; iii) if possible, inositol hydroxyls should form H-bond contacts with nearby H-bond acceptors; iv) while bound to the enzyme, the acyl chains attached to the glycerol backbone of the acceptor should be free to insert into the cell membrane. The validity of the resulting model of the ternary complex was tested by mutagenesis.

RESULTS

Structure determination. Crystals of *C. glutamicum* PimB' were grown in two crystal forms with triclinic and orthorhombic lattices, depending on whether the recombinant protein carried a His_6-affinity tag at the N- or C-terminus, respectively. Incubation of N-terminally tagged PimB' with the native donor substrate guanosine-5'-diphosphate-α-D-mannopyranose (GDP-Man, 10 mM) produced triclinic crystals, based on which the structure was determined (see Experimental Procedures, supplemental Fig. 2S). The refined 'triclinic' model (2.2 Å resolution, Fig. 2*A*, Table 1) comprises two molecules of PimB' (residues 4 – 381) in the asymmetric unit, missing only the first three residues of *C. glutamicum* PimB'. The two copies were refined applying tight NCS-restraints (root mean square deviation (RMSD) between 378 Cα pairs of 0.03 Å, estimated coordinate error 0.37 Å, Table 1). Subsequent to the structure determination, an orthorhombic crystal form was obtained and refined to a resolution of 2.0 Å. The lattices of the orthorhombic and triclinic crystal forms are similar (triclinic a_t = 44 Å, b_t = 50 Å, c_t = 85 Å; $\alpha \approx \beta \approx \gamma \approx 90^o$; orthorhombic: $a_o = a_t$, $b_o = c_t$, $c_o \approx 2\ b_t$, Table 1), but show distinct packing interfaces. This is presumably due to the affinity tag, which is disordered in the orthorhombic lattice, but partially ordered in the triclinic crystal form. Weak density for GDP-Man was observed in both copies of PimB' in the 'triclinic' model, suggesting partial occupancy. In contrast, the GDP- and GDP-Man-bound structures of the orthorhombic form show strong density for the nucleotide, although the mannose moiety is disordered (see below).

Overall structure of PimB'. The structure of *C. glutamicum* PimB' displays the canonical fold of GT-B glycosyltransferases, comprising two Rossmann fold-like domains (41) that are connected by a hinge region with the substrate binding sites located in the cleft between the domains (Fig. 2*A*). The N-terminal domain (residues 4 – 171, 364 – 381), assumed to bind the acceptor substrate, is composed of a 7-stranded, parallel β-sheet (strands β1 - β7), with connecting helices α1 to α6. The C-terminal domain (residues 172 – 363) binds the nucleotide-sugar donor substrate, and consists of a 6-stranded, parallel β-sheet (strands β8 - β13) with connecting helices α8 to α11, and $3_{10}6$. A conserved feature of GT4 family enzymes, and indeed of the GT-B fold, are the consecutive C-terminal helices α13 and α14 that span the length of the protein and belong to the C- and N-terminal domains, respectively. Domain reorientation is facilitated by a flexible hinge between the domains, consisting of the β7-α7 loop, with invariant Gly171 at its centre, and the short linker separating helices α13 and α14, with highly conserved Trp365 clamped between the domains (Fig. 2*A*, supplemental Fig. 1S).

Differences in crystal packing between the orthorhombic and triclinic crystal forms resulted in a slightly different inter-domain orientation (4º rotation), and an additional helix, α5' (residues 120 – 125), in the active site cleft (Fig. 3*A*). In the domain-wise superposition, the triclinic and orthorhombic models match closely (RMSD of 1.27 Å for N-terminal domain, 0.39 Å for the C-terminal domain), but individual residues in the region of helix α5' undergo considerable shifts of their position. For instance, the α-carbon of Trp124 is shifted by 7 Å, and its side chain switches from contacts with the C-terminal domain (Gly285, Leu287 of the β11-α11 loop) to folding back onto the N-terminal domain in the orthorhombic form.

Structural neighbors. We searched for structural homologs of PimB' in the PDB through distance matrix alignment (DALI, (42)). The closest neighbor with respect to overall structure was glycosyltransferase MshA from *C. glutamicum* (Fig. 2*B*, PDB entry 3c48, Z = 23.7, RMSD 1.6 Å, 180 aligned Cα pairs of 393 residues, sequence identity 31%) (29), followed by glycosyltransferases *E. coli* WaaG (2iw1, Z 21.8, 2.3 Å, 180/370 Cα pairs, 25% id.) (26), and *Bacillus anthracis* Ba1558 (2jjm, Z 21.3, 1.8 Å, 177/359 Cα pairs, 21% id.) (28). The fourth hit was the functionally related α-mannosyltransferase *M. smegmatis* PimA (Fig. 2*C*, 2gej, Z 20.1, 1.9 Å 176/361 Cα pairs, 24% id.). Restricting the DALI search to the more variable N-terminal domain, which commonly binds the acceptor substrate, the closest neighbor was the N-terminal domain of *B. anthracis* Ba1558 (2jjm, Z 19.2, 2.1 Å, 165 Cα, 18% id), followed by those of MshA (3c48, Z 16, 2.6 Å, 172 Cα, 10% id.) and PimA (2gej, Z 15.4, 161 Cα, 17% id.).

The structural comparison between PimB' and PimA (Fig. 2*C*) is of particular interest as they catalyze subsequent steps in PIM biosynthesis utilizing identical saccharide donors (GDP-Man), yet showing distinct regio-selectivity for mannosylation of the inositol moiety of PI. The domain-wise superposition of PimB' and PimA shows conserved cores for either domain (N-terminal domain: RMSD 1.53 Å for 106 Cα pairs; C-terminal domain: RMSD 1.19 Å for 147 Cα pairs). Yet, several structural differences are noteworthy. First, PimB' lacks the extended β3-α3 loop, due to a 10-residue deletion (supplemental Fig. 1S). This loop, referred to as 'IP-binding' loop (10, 29), has been implicated in acceptor substrate binding in both PimA and MshA. While fully ordered in PimB', it is disordered in nucleotide-bound PimA (Fig. 2*C*). Second, the β11-α11 loop in PimB' has a 6-residue insertion relative to PimA. This loop makes contacts with the structurally variable loop β5–α5 (containing helix α5' in orthorhombic PimB'). Furthermore, PimB' comprises additional helices in the N-terminal (α2) and C-terminal domains (α7, Fig. 2*C*), while helix α5 of PimB' overlaps partially with helices α4 and α5 in PimA.

Active site and nucleotide binding. We crystallized PimB' in the presence of GDP or GDP-Man, which bind with affinities in the order of 20 – 40 μM in terms of K_d (Table 2). Density for the nucleotide is weak in the triclinic crystal form (10 mM GDP-Man present during crystallization, Fig. 4*A*), but is well defined for the orthorhombic crystals (20 mM GDP or GDP-Man, Fig. 4*B* and *C*). The nucleotide binding site is situated in the cleft between N- and C-terminal domains of PimB' (Fig. 3*B*), a conserved feature of GT-B glycosyltransferases. In all cases, binding of the nucleotide (GDP or GDP-Man) resulted in crystallization of an 'open' configuration of PimB', in marked contrast to the donor substrate-bound structures of PimA and MshA (supplemental Fig. 3S). In this configuration, the central cleft is open and the nucleotide makes no contacts with residues in the N-terminal domain. A small, but significant density blob in triclinic PimB' indicates the presence of mannose (Fig. 4*A*), but the blob fails to define the conformation of the saccharide moiety unequivocally. Paradoxically, the mannose moiety is disordered in the orthorhombic crystals of GDP-Man-bound PimB' (Fig. 4*B*), although the ligand density is much stronger overall. We attribute the absence of mannose density to the absence of stabilizing contacts in the open form of the enzyme. We found no indication of mannose-release in the absence of acceptor substrate when incubating PimB' with [^{14}C]-labeled GDP-Man over a period of several hours.

With few exceptions, residues in contact with the nucleotide reside in the following structural elements: the β8-α8 loop, the β10-α10 loop and the β11-α11 loop (Fig. 3*B*). The purine base forms hydrogen bonds with the amide nitrogen and carbonyl groups of Leu261

(β10-α10 loop), but no specificity determining contacts with side chain atoms. The side chains of Val235 (C-terminal end of β9), Arg260 and Met266 (β10-α10 loop), contact the base by way of apolar and van der Waals contacts. Interestingly, in the (low-occupancy) 'triclinic' structure, Arg260 is only partially ordered. The C2' and C3' hydroxyls of the ribose make conserved H-bond interactions with the side chain carboxyl of Glu298 (helix α11). The pyrophosphate group is set against a surface of positive electrostatic potential of residues Arg206, Arg210 and Lys211 (β8-α8 loop), and the 'oxyanion hole'-like surface of four consecutive amide nitrogen groups (residues 291 to 294) in the β11-α11 loop. The carboxylate group of Glu290, a highly conserved residue in the β11-α11 loop, is held in a fixed orientation through ionic bonds with Lys211 and Arg210, neutralizing its negative charge that would otherwise repel the close-by β-phosphate (d ~ 3.8 Å).

It was surprising to find the pyrophosphate in two distinct conformations that differ with respect to the position of the α-phosphate, whereas the position of the β-phosphate is more or less fixed, stabilized by ionic contacts with the guanido group of Arg206 and the terminal ammonium group Lys211 (Fig. 3*B* and 4). Recording diffraction data of three independent GDP-bound complexes, the pyrophosphate of GDP consistently displayed a conformation that coincides with that seen in the UDP-bound complex of trehalose-6-phosphate synthase (*E. coli* OtsA, PDB entries 1uqu, 2wtx (43-45)). In this 'OtsA-like' conformation, the α-phosphate is positioned close to the main chain amide nitrogen groups of Ile294 and Val295 (N-terminal of helix α11), forming weak H-bonds (3.3 to 3.5 Å) (Fig. 4*C*). In contrast, orthorhombic crystals of PimB' in complex with GDP-Man show a superposition of the 'OtsA-like' conformer and a second conformation that resembles the nucleotide conformation in GDP-Man-bound PimA (Fig. 4*B*) (2gek, 2gej) (27). In the PimA-like conformation, the α-phosphate is positioned within H-bond distance (2.9 Å) of Nε of Arg206. The conversion between the OtsA- and PimA-like conformers occurs through a rotation of approximately ~120° about the C4'-C5' bond of the ribose, drastically altering the position of the α-phosphate and shifting the β-phosphate by about 0.5 – 0.6 Å away from Arg206, while the guanosine moiety remains stationary. Interestingly, the triclinic crystal form shows density only for the PimA-like conformer (Fig. 4*A*).

Modeling the ternary complex. In order to understand the structural basis of regio-selectivity of mannosyl transfer, we modeled the closed configuration of PimB', as this is assumed to represent the catalytically competent conformation. Attempts to crystallize ternary complexes of PimB' bound to GDP and acceptor substrate analogs were not successful. Crystals formed readily when (C-terminally tagged) PimB' was incubated with excess amounts of GDP and of either D-*myo*-inositol-1-phosphate or deacylated PIM_1. Nonetheless, the diffraction data consistently revealed density for the nucleotide, but none for the acceptor. Binding studies for PimA (27) and kinetic studies of MshA (46) showed that the substrates bind sequentially, with the sugar-nucleotide binding first, followed by the acceptor. This scenario very likely applies to PimB' as well, and acceptor binding may require the 'closed' configuration, which, in the case of PimB', appears to be disfavored in crystal lattice formation.

The model of the closed form of the enzyme was constructed using the template of the ternary complex of *C. glutamicum* MshA (PDB entry 3c4v), the closest structural neighbor of PimB' according to our DALI search. The structures of PimB' (orthorhombic form) and MshA were aligned by secondary structure matching with respect to their C-terminal domains, followed by aligning the N-terminal domain of PimB' to the N-terminal domain of MshA, which involves a 27°-rotation (Fig. 5*A*). The rigid body rotary movement of the N-terminal domain towards the C-terminal domain leads to minor steric clashes between helix α5' and the β11-α11 loop. However, as the comparison between the 'triclinic' and 'orthorhombic' structures of PimB' has demonstrated, the α5' region is conformationally flexible. Therefore, it is justified to assume that this region will adapt to spatial constraints imposed by the rotary movement.

As density for the mannose moiety was not sufficiently defined, we modeled its orientation according to the nucleotide-on-nucleotide superposition of PimB' with the UDP-glucose bound structure of OtsA. The latter complex

structure in conjunction with the ternary complex of OtsA bound to UDP and a disaccharide mimetic inhibitor has provided the most compelling structural evidence as yet to support the proposed S_Ni internal return mechanism for retaining glycosyltransferases of the GT-B fold (see Discussion) (24, 43-45).

Despite restricting the protein modeling to a rigid body motion of the N-terminal domain and positioning the mannose moiety of GDP-Man according to a structural superposition between enzymes that share only 15% sequence identity, the resulting structural model displays a number of mechanistically plausible features. First, the rotary motion places residue Ile18, which precedes the Gly-Gly motif, on the face of the guanosine base, analogous to the interaction of the corresponding residues in PimA (Pro), and OtsA (Ala) (Fig. 5*B*). The resulting closed binding cavity for the sugar-nucleotide can accommodate GDP-Man without steric overlap. The binding cavity also allows the nucleotide to adopt both conformations that we observed for the pyrophosphate. Importantly, the binding cavity shows a prominent opening, which selectively exposes the C2 hydroxyl and the C1 carbon of GDP-mannose to the putative acceptor-binding site (Fig. 5*B*).

We manually docked a model of PIM_1 (2-mannosyl-phosphatidyl-*myo*-inositol-1-phosphate) into the cleft of the 'closed' enzyme (shown in cyan in Fig. 5), such that the 6-hydroxyl is pointing towards the exposed mannose moiety, while the mannosylated C2 is pointing in the opposite direction. In this way, the 3- and 4-OH groups of the inositol ring are positioned within H-bonding distance to the carboxylate of Asp13 in the β1-α1 loop, providing a bidentate anchoring interaction that defines the orientation of the inositol ring. In PimA, the residue corresponding to Asp13 is Tyr9, which has been previously implicated in acceptor substrate specificity of mannosyl transfer (10, 14, 27). The phosphate moiety is positioned within 3.5 Å of the guanido group of Arg210, suggesting that Arg210 might contribute to acceptor recognition. We also note that in this model, the position of the phosphate moiety falls in close range of an ordered sulfate molecule of the triclinic structure, which is coordinated by Arg206 (Fig. 3*B*). The model of the closed complex then served as a guide to probe the structural determinants of acceptor binding by site-directed mutagenesis.

Mutagenesis. In order to assess the significance of selected residues for PimB' function, we probed enzyme activity by monitoring the transfer of [^{14}C]-labeled mannose from GDP-Man to Ac_1PIM_1 (Fig. 6*A*, Table 2). Products were analyzed in terms of radioactivity and by thin layer chromatography (Fig. 6*A*), while nucleotide binding was assayed by intrinsic tryptophan fluorescence (Fig. 6*B*). In addition, proper folding of mutants was ascertained by circular dichroism spectroscopy (Fig. 6*C* and supplemental Fig. 4S).

Residues of the C-terminal domain that contact the nucleotide, very sensitively influence enzymatic activity. Residual activity of single-residue mutants of the three basic side chains (Arg206, Arg210 and Lys211) situated in the vicinity of the pyrophosphate is less than 9% that of wild-type (Fig. 3*B*, Table 2). Likewise, mutation of the conserved Glu residue at position 290 to either Gln or Asp reduced activity by more than 95% (Fig. 6*A*, Table 2). While detrimental for activity, all of these mutants still bind the acceptor substrate with appreciable affinity. Thus, these mutations generally weaken, but do not abrogate binding of GDP or GDP-Man.

While the nucleotide binding site is highly conserved between PimA and PimB', subtle structural differences are evident. For instance, Gly291 in PimB' corresponds to a serine in PimA. Modeling the GDP-Man in PimB' according to the PimA:GDP-Man complex indicated potential steric hindrance between a serine on position 291 and the mannose moiety. Indeed, compared to wild-type PimB', the G291S mutant shows 95% reduced activity, and a nearly 4-fold increase in K_d for binding of GDP-Man (Table 2). In contrast, the substitution of Ser205 with the corresponding Gly in PimA increases activity by about 10%, although the increase is within one standard deviation.

DISCUSSION

PIM biosynthesis occurs through a series of enzyme-catalyzed reactions during which PI is decorated on the inositol moiety with mannosyl groups in an ordered and regio-selective manner. The present structure in combination with the mutational analysis sheds light on how regio-selectivity is achieved and provides indirect evidence for the hypothesis that

catalysis occurs through a substrate-mediated S_Ni internal return mechanism.

In terms of fold and overall structure, PimB' conforms to the paradigm of GT-B glycosyltransferases. However, a couple of observations are noteworthy. First, nucleotide binding did not result in crystallization of a closed configuration of the enzyme, in contrast to the closed states observed for the nucleotide-bound complexes of PimA, MshA, OtsA or WaaG (26, 27, 29, 45). However, it seems improbable that PimB' is unable to close the active site cleft, given the close similarity to the structural homologs. Small changes in orientation of the domains between the orthorhombic and triclinic crystal form, provide direct evidence for conformational flexibility. More importantly, the G20W mutation leads to 97% drop in activity, but binds the donor with affinity similar to that of wild type PimB' (Table 2). The fact that a bulky side chain in the centre of the cleft is tolerated in terms of nucleotide binding, but abrogates activity is consistent with the notion that it impedes the closing motion.

Second, the conformational flexibility observed in the β5-α5 loop of the active site cleft has not been observed previously in GT4 glycosyltransferases. The shift of individual side chains and the corresponding change of intramolecular contacts seems significant. For instance, the position of Trp124's Cα is shifted by 7 Å, and its side chain makes contacts with different parts of the structure in the two crystal forms. We probed whether helix α5' is required for activity by substituting glycine at position 123 with proline, but the resulting drop of activity is only about 15% (Table 2). Furthermore, His120 at the N-terminus of helix α5' is only ordered in the triclinic structure and disordered when helix α5' forms. Substituting His120 by serine reduced activity by about 20% (Table 2), while the CD spectrum of the mutant G123P is virtually identical to that of wild-type PimB' (supplemental Fig. 4S). In light of this data, we hypothesize that the conformational flexibility of the β5-α5 loop works in concert with the open/closing motion of the active site cleft upon substrate binding, but influences catalysis only in an indirect fashion, if at all.

PimB' is a GT-B glycosyltransferase that retains the stereochemical configuration of the anomeric carbon of the donor saccharide. The catalytic reaction mechanism of retaining GT-B transferases has remained somewhat of a mystery. In glycoside hydrolases, retention of the anomeric configuration occurs through an S_N2-like double displacement mechanism, mediated by a pair of side chain carboxylates and involving a covalent sugar-enzyme intermediate (47). However, none of the inhibitor- or substrate mimic-bound complexes solved to date has provided structural evidence to support such a scenario in GT-B glycosyltransferases (24, 48-50). Instead, a substrate-mediated S_Ni (internal return) mechanism has been invoked, reviewed in reference (24). This mechanism involves three tightly linked reaction steps (Scheme 1): first, decay of the sugar-phosphate bond, leaving a (solvent-separated) ion pair of the oxocarbenium ion-like saccharide and the negatively charged phosphate (panels 2,3); second, deprotonation of the acceptor hydroxyl by the phosphate leaving group; third, nucleophilic attack by the acceptor on the oxocarbenium ion (panels 3,4). The first step is rate limiting, while the subsequent steps must occur on a time scale shorter than solvent attack or ion pair re-organization (24). If correct, this mechanism imposes constraints on the geometry of donor and acceptor substrate in the ternary complex, in that the leaving group must be sufficiently close to the acceptor to be able to deprotonate the attacking acceptor hydroxyl.

In our model of the acceptor-bound complex, the 6-OH of inositol (O_i6) is positioned within ~ 4 Å of the C1 carbon of the donor mannose, and within 5 Å of the β-phosphate (Pβ). The distance vectors O_i6–C1 and C1–Pβ enclose an angle of ~ 105º. Thus, the geometric configurations of the substrates in the ternary complex model are compatible with the requirements of the S_Ni reaction mechanism. Our mutagenesis data provide further supporting evidence for a substrate-dependent reaction mechanism. For instance, the most obvious candidate to act as catalytic nucleophile in an S_N2-like mechanism is Glu290, a carboxylate that is strictly conserved across GT4 family enzymes, and positioned close (~ 4.5 Å) to the sugar-phosphate bond. Switching this residue to glutamine, which should completely impair nucleophilic attack, decreases activity by about 95%, while leaving substrate binding largely intact (< 3-fold increase of K_d). Importantly, the residual activity is about 10 times above baseline level, and can be reduced further by the E290D substitution (Table 2). In contrast, the

substitution of highly conserved His118 by serine decreases activity essentially to baseline level. In the 'open' configuration, H118 does not contact the donor substrate, but in the 'closed' model contacts the C6 hydroxyl of the nucleotide-linked mannose. His118 is positioned closely opposite to Ile21 (closest contact 3.9 Å). Substitution of Ile21 with either Ser or Ala diminishes activity to about 2.5%. The close proximity of a hydrophobic side chain to His118 suggests that the latter is pushed to its non-ionized state, which could enable the His side chain to act as a nucleophile. However, in our modeled ternary complex as in the experimental structures of OtsA (2wtx) or MshA (3c4v), the position of His118 relative to the nucleotide makes such a role improbable.

In their structural evaluation of PimA, Guerin and colleagues (27) identified Tyr9 as a residue that might form a stacking interaction with the inositol ring of PI, and thus may be key to acceptor specificity. This tyrosine is located in the β1-α1 loop and corresponds to Asp13 in PimB'. Within the sequences of PimB' and PimA enzymes, respectively, the Asp and Tyr are strictly conserved at this position (supplemental Fig. 1S). In the model of the ternary complex, Asp13 is positioned within H-bond distance to the 3- and 4-OH of the inositol moiety, *i.e.* the only inositol hydroxyls that are in an equatorial configuration. Such a bidentate interaction would ensure that inositol presents the axial 6-OH, rather than the 2-OH to the C1 carbon of mannose. Indeed, swapping the positions of 3-OH and 4-OH of inositol would present the 5-OH to C1 of mannose, rather than 2-OH. Probing Asp13 by mutagenesis, it is intriguing that even the most conservative substitution - aspartate to asparagine - reduces activity by 98.5% relative to wild type. Indeed, this subtle change knocks down activity as effectively as the more drastic Asp to Ala substitution (Table 2). Residues in the immediate vicinity – Asn12, Ile21 and Gln22 – are similarly sensitive to substitutions, causing a reduction of activity by at least 70%. We also probed whether the specificity of PimB' can be switched to mannosylation of the 2-OH by substituting Asp13 by tyrosine. However, the result was a knock-down of activity to blank level (Table 2). Thus, residues in the β1-α1 loop play a key role in acceptor recognition, but the single substitution on position 13 is not sufficient to switch acceptor specificity.

The deleterious effect of the D13N mutation stresses the significance of Asp13 in determining acceptor specificity, although the mechanistic implication of this substitution is not clear, as an asparagine should still be able to form H-bonds with the inositol hydroxyl groups. However, the effect is reminiscent of an analogous mutation in glycosidase family GH68. There, substitution of a strictly conserved Asp by Asn, which formed strong H-bonds with saccharide substrate hydroxyls, but, for geometric reasons, could not be a nucleophile or general acid/base, reduced activity by 2 orders of magnitude in terms of k_{cat}. This effect was attributed to stabilization of the oxocarbenium ion transition state by the negatively charged carboxylate (51, 52). Lack of sequence conservation across GT4 family enzymes appears to rule out a fundamental role of Asp13 in a general reaction mechanism of GT4 enzymes. However, we note that PimA extends a glutamic acid side chain (residue 199) towards the inositol ring when superimposed with our model of the PimB' ternary complex. This glutamic acid, which is critical for activity (10), is located in the β8-α8 loop, and, in the structural superposition, lines up with Val208 of PimB'. Our model suggests that Glu199 could form H-bond contacts with the equatorial OH groups of inositol, if the inositol moiety is oriented such that the 2-OH is presented to the C1 of the donor saccharide. Thus, the active site cleft of PimA includes a side chain that could play a role functionally analogous to that of Asp13 in PimB'.

The dual conformation of the pyrophosphate in complexes co-crystallized with GDP-Man is intriguing. This feature could be mechanistically linked to release of the donor mannosyl group following decay of the sugar-phosphate bond and transfer to inositol. PimB':GDP complexes have a pronounced preference for the 'OtsA'-like nucleotide conformation, while GDP-Man bound PimB' either displayed the superposition of the OtsA- and PimA-like states or only the PimA-like state. Density for the mannose group is completely missing in the orthorhombic complex. Yet the presence of a disordered sugar moiety can be inferred on the one hand from the pronounced preference for a dual nucleotide conformation in the GDP-Man co-crystals and, secondly, from the positioning of an extra water molecule between β-phosphate and Glu290 in

the GDP co-crystals (W4, Fig. 4*C*). These observations were consistent between at least two different crystals of each complex. Between the PimA- and OtsA-like conformations, the β-phosphate shifts away from Arg206 by about 0.5-0.6 Å, weakening the corresponding ionic interaction, which would facilitate increased mobility of Arg206. Such increased mobility of this side chain could promote release of the transferred mannosyl group from the confines of the donor substrate-binding cavity.

Concluding remarks. Biosynthesis of PIM occurs through a series of glycosyltransfer reactions that initially take place at the cytoplasmic side of the membrane. The final product is presented on the extracellular face of the membrane, and hence flippase-mediated membrane translocation of either the final or an intermediate product must occur along the way (3). We consider the following scenario probable: PimA, PimB' and PimC act on acceptor substrates - PI and intermediate forms of PIM - that are anchored in the cytoplasmic leaflet of the membrane. The active site geometry of PimB' and its electrostatic surface potential (supplemental Fig. 5S) are compatible with the notion of transient membrane association, potentially in the orientation shown in Fig. 5A. Liposome binding data recorded for PimA further support this scenario (27). PimB' differs from PimA in several structural details that reflect their distinct acceptor substrate specificities. Chief among them is the lack of the extended β3-α3 loop, which contributes to acceptor binding in PimA, but, based on geometry, cannot play such a role in PimB'. Second, the active site cleft shows a more open surface, to accommodate the mannose attached to the 2-position of inositol. Third, we show that Asp13, which is critical for activity, cannot be swapped for the corresponding tyrosine in PimA. We postulate that Asp13 is chiefly responsible for regio-specific orientation of inositol relative to the donor saccharide. Our mutagenesis data also lend indirect support to the hypothesis that glycosyltransfer with retention of the α-configuration of the anomeric carbon occurs through a substrate-mediated S_Ni-type mechanism.

REFERENCES

1. Jozefowski, S., Sobota, A. and Kwiatkowska, K. (2008) *Bioessays* **30**, 943-954
2. Guenin-Mace, L., Simeone, R. and Demangel, C. (2009) *Transbound Emerg Dis* **56**, 255-268
3. Berg, S., Kaur, D., Jackson, M. and Brennan, P. J. (2007) *Glycobiology* **17**, 35-56R
4. Doz, E., Rose, S., Court, N., Front, S., Vasseur, V., Charron, S., Gilleron, M., Puzo, G., Fremaux, I., Delneste, Y., Erard, F., Ryffel, B., Martin, O. R. and Quesniaux, V. F. (2009) *J Biol Chem* **284**, 23187-23196
5. Rhoades, E. R., Archambault, A. S., Greendyke, R., Hsu, F. F., Streeter, C. and Byrd, T. F. (2009) *J Immunol* **183**, 1997-2007
6. Fischer, K., Scotet, E., Niemeyer, M., Koebernick, H., Zerrahn, J., Maillet, S., Hurwitz, R., Kursar, M., Bonneville, M., Kaufmann, S. H. and Schaible, U. E. (2004) *Proc Natl Acad Sci U S A* **101**, 10685-10690
7. de la Salle, H., Mariotti, S., Angenieux, C., Gilleron, M., Garcia-Alles, L. F., Malm, D., Berg, T., Paoletti, S., Maitre, B., Mourey, L., Salamero, J., Cazenave, J. P., Hanau, D., Mori, L., Puzo, G. and De Libero, G. (2005) *Science* **310**, 1321-1324
8. Driessen, N. N., Ummels, R., Maaskant, J. J., Gurcha, S. S., Besra, G. S., Ainge, G. D., Larsen, D. S., Painter, G. F., Vandenbroucke-Grauls, C. M., Geurtsen, J. and Appelmelk, B. J. (2009) *Infect Immun* **77**, 4538-4547
9. Mahon, R. N., Rojas, R. E., Fulton, S. A., Franko, J. L., Harding, C. V. and Boom, W. H. (2009) *Infect Immun* **77**, 4574-4583
10. Guerin, M. E., Schaeffer, F., Chaffotte, A., Gest, P., Giganti, D., Kordulakova, J., van der Woerd, M., Jackson, M. and Alzari, P. M. (2009) *J Biol Chem* **284**, 21613-21625

11. Kordulakova, J., Gilleron, M., Mikusova, K., Puzo, G., Brennan, P. J., Gicquel, B. and Jackson, M. (2002) *J Biol Chem* **277**, 31335-31344

12. Kordulakova, J., Gilleron, M., Puzo, G., Brennan, P. J., Gicquel, B., Mikusova, K. and Jackson, M. (2003) *J Biol Chem* **278**, 36285-36295

13. Lea-Smith, D. J., Martin, K. L., Pyke, J. S., Tull, D., McConville, M. J., Coppel, R. L. and Crellin, P. K. (2008) *J Biol Chem* **283**, 6773-6782

14. Guerin, M. E., Kaur, D., Somashekar, B. S., Gibbs, S., Gest, P., Chatterjee, D., Brennan, P. J. and Jackson, M. (2009) *J Biol Chem* **284**, 25687-25696

15. Mishra, A. K., Batt, S., Krumbach, K., Eggeling, L. and Besra, G. S. (2009) *J Bacteriol* **191**, 4465-4472

16. Schaeffer, M. L., Khoo, K. H., Besra, G. S., Chatterjee, D., Brennan, P. J., Belisle, J. T. and Inamine, J. M. (1999) *J Biol Chem* **274**, 31625-31631

17. Tatituri, R. V., Illarionov, P. A., Dover, L. G., Nigou, J., Gilleron, M., Hitchen, P., Krumbach, K., Morris, H. R., Spencer, N., Dell, A., Eggeling, L. and Besra, G. S. (2007) *J Biol Chem* **282**, 4561-4572

18. Mishra, A. K., Klein, C., Gurcha, S. S., Alderwick, L. J., Babu, P., Hitchen, P. G., Morris, H. R., Dell, A., Besra, G. S. and Eggeling, L. (2008) *Antonie Van Leeuwenhoek* **94**, 277-287

19. Kremer, L., Gurcha, S. S., Bifani, P., Hitchen, P. G., Baulard, A., Morris, H. R., Dell, A., Brennan, P. J. and Besra, G. S. (2002) *Biochem J* **363**, 437-447

20. Morita, Y. S., Sena, C. B., Waller, R. F., Kurokawa, K., Sernee, M. F., Nakatani, F., Haites, R. E., Billman-Jacobe, H., McConville, M. J., Maeda, Y. and Kinoshita, T. (2006) *J Biol Chem* **281**, 25143-25155

21. Morita, Y. S., Patterson, J. H., Billman-Jacobe, H. and McConville, M. J. (2004) *Biochem J* **378**, 589-597

22. Patterson, J. H., Waller, R. F., Jeevarajah, D., Billman-Jacobe, H. and McConville, M. J. (2003) *Biochem J* **372**, 77-86

23. Liu, J. and Mushegian, A. (2003) *Protein Sci* **12**, 1418-1431

24. Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G. (2008) *Annu Rev Biochem* **77**, 521-555

25. Coutinho, P. M., Deleury, E., Davies, G. J. and Henrissat, B. (2003) *J Mol Biol* **328**, 307-317

26. Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2006) *Chem Biol* **13**, 1143-1152

27. Guerin, M. E., Kordulakova, J., Schaeffer, F., Svetlikova, Z., Buschiazzo, A., Giganti, D., Gicquel, B., Mikusova, K., Jackson, M. and Alzari, P. M. (2007) *J Biol Chem* **282**, 20705-20714

28. Ruane, K. M., Davies, G. J. and Martinez-Fleites, C. (2008) *Proteins* **73**, 784-787

29. Vetting, M. W., Frantom, P. A. and Blanchard, J. S. (2008) *J Biol Chem* **283**, 15834-15844

30. Chua, T. K., Bujnicki, J. M., Tan, T. C., Huynh, F., Patel, B. K. and Sivaraman, J. (2008) *Plant Cell* **20**, 1059-1072

31. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) *J Appl Crystallogr* **40**, 658-674

32. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26**, 795-800

33. Sheldrick, G. M. (2008) *Acta Crystallogr A* **64**, 112-122

34. CCP4 (1994) *Acta Crystallogr D Biol Crystallogr* **50**, 760-763

35. de la Fortelle, E. and Bricogne, G. (1997) *Methods Enzymol* **276**, 472-494

36. Collier, A. and Wagner, G. K. (2008) *Chem Commun (Camb)* 178-180

37. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A. (2007) *Curr Protoc Protein Sci* **Chapter 2**, Unit 2.9

38. Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**, 240-255

39. Dobson, G., Minnikin, D. E., Minnikin, S. M., Parlett, J. H., Goodfellow, M. and Magnusson, M. (1985) in *Chemical methods in bacterial systematics* (Goodfelow, M. and Minnikin, D. E., eds.) pp. 237-265, Academic Press, London

40. Besra, G. S., Morehouse, C. B., Rittner, C. M., Waechter, C. J. and Brennan, P. J. (1997) *J Biol Chem* **272**, 18460-18466

41. Rao, S. T. and Rossmann, M. G. (1973) *J Mol Biol* **76**, 241-256

42. Holm, L. and Sander, C. (1996) *Science* **273**, 595-603

43. Errey, J. C., Lee, S. S., Gibson, R. P., Martinez Fleites, C., Barry, C. S., Jung, P. M., O'Sullivan, A. C., Davis, B. G. and Davies, G. J. (2010) *Angew Chem Int Ed Engl* **49**, 1234-1237

44. Gibson, R. P., Tarling, C. A., Roberts, S., Withers, S. G. and Davies, G. J. (2004) *J Biol Chem* **279**, 1950-1955

45. Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R. and Davies, G. J. (2002) *Chem Biol* **9**, 1337-1346

46. Newton, G. L., Ta, P., Bzymek, K. P. and Fahey, R. C. (2006) *J Biol Chem* **281**, 33910-33920

47. Zechel, D. L. and Withers, S. G. (2001) *Curr Opin Chem Biol* **5**, 643-649

48. Sheng, F., Jia, X., Yep, A., Preiss, J. and Geiger, J. H. (2009) *J Biol Chem* **284**, 17796-17807

49. Steiner, K., Hagelueken, G., Messner, P., Schaffer, C. and Naismith, J. H. (2010) *J Mol Biol* **397**, 436-447

50. Frantom, P. A., Coward, J. K. and Blanchard, J. S. (2010) *J Am Chem Soc* **132**, 6626-6627

51. Meng, G. and Futterer, K. (2003) *Nat Struct Biol* **10**, 935-941

52. Yanase, H., Maeda, M., Hagiwara, E., Yagi, H., Taniguchi, K. and Okamoto, K. (2002) *J Biochem* **132**, 565-572

53. Engh, R. and Huber, R. (1991) *Acta Crystallogr A* **47**, 392-400

54. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. r., Snoeyink, J., Richardson, J. S. and Richardson, D. C. (2007) *Nucleic Acids Res* **35**, W375-83

FOOTNOTES

GSB acknowledges support in the form of a Personal Research Chair from Mr. James Bardrick, Royal Society Wolfson Research Merit Award, as a former Lister Institute-Jenner Research Fellow. GSB and KF acknowledge support from The Wellcome Trust (081569/Z/06/Z).

The atomic coordinates and structure factors (codes 3oka, 3okc and 3okp) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

The abbreviations used are: TB, tuberculosis; PI, phosphatidyl-*myo*-inositol-1-phosphate; GDP-Man, guanosine-5'-diphosphate-α-D-mannose; LM, lipomannan; LAM, lipoarabinomannan; PIM, phosphatidylinositol mannoside; AcPIM, acylated phosphatidylinositol mannoside; GT-A,

glycosyltransferase fold type A; GT-B, glycosyltransferase fold type B; GT4, glycosyltransferase family 4; SAD, single-wavelength anomalous dispersion; MIR, multiple isomorphous replacement; NCS, non-crystallographic symmetry; MR, molecular replacement; TLC, thin layer chromatography; MALDI-TOF-MS, matrix-assisted laser desorption time-of-flight mass spectrometry; RMSD, root mean square deviation.

FIGURE LEGENDS

Fig. 1 Schematic diagram of phosphatidylinositol-mannosides (PIM) and pathway of PIM synthesis. ManT and AraT designate generic mannosyl- and arabinosyl transferase enzymes, respectively. LM = lipomannan, LAM = lipoarabinomannan.

Fig. 2 Ribbon diagram of *C. glutamicum* PimB' in the triclinic crystal form and comparison with PimA and MshA. *A,* N-terminal domain (blue) and C-terminal domain (green) are shown together with a stick model of the bound nucleotide. *B,* Comparison with *C. glutamicum* MshA (3c4v, in beige), after superposition with respect to secondary structure elements in the C-terminal domain. *C,* Comparison with the structure of PimA (beige) following domain-wise superposition onto PimB'. Beige spheres indicate the ends of the disordered 'IP-binding' loop of PimA.

Fig. 3 Structural flexibility in the β5-α5 loop and overview of the active site of PimB'. *A,* Comparison of conformation between the triclinic (green/blue) and orthorhombic crystal (grey) forms of PimB'. Helix α5' is only present in the orthorhombic form. *B,* Close-up view of the active site cleft of triclinic PimB' in the open configuration with ribbons in blue and green indicating the N- and C-terminal domains, respectively. Loop regions involved in nucleotide binding are coloured in magenta. Residues in contact with nucleotide or that were probed by point mutagenesis are shown in yellow sticks. The nucleotide shown reflects the 'PimA'-like conformation of the pyrophosphate. The mannose moiety is shown in transparent sticks, modeled according to the conformation seen in OtsA.

Fig. 4 Different modes of nucleotide binding in active site of PimB'. *A,* σ_A-weighted Fo – Fc density map (contour levels 2.2 σ (blue) and –2.2 σ (red)) of triclinic PimB'. The green circle marks the position of the density blob corresponding to the GDP-linked mannose. *B,* and *C,* σ_A-weighted Fo – Fc density maps (+2.8 σ (blue), -2.8 σ (red)) of GDP-Man (in *B*) and GDP-bound (in *C*) complexes of orthorhombic PimB'. Selected water molecules are labelled in blue and dashed lines in magenta indicate a subset of relevant non-covalent interactions. All maps were calculated with Fc amplitudes and phases of the ligand-free model.

Fig. 5 Model of the closed form of PimB' with docked acceptor. *A,* Ribbon diagram of the closed configuration of PimB' superimposed with the molecular surface of the open structure of GDP-Man-bound PimB' (orthorhombic crystal form). The dashed line (black) indicates the approximate position of the membrane surface. *B,* Close-up view of the docked ternary complex. PIM_1 (cyan bonds) is shown without the acyl chains and the glycerol backbone. Side chains probed by mutagenesis are show in blue (N-terminal domain) and green (C-terminal domain).

Fig. 6 Activity and nucleotide binding of *C. glutamicum* PimB' and of selected point mutants. *A,* Analysis of transfer of [^{14}C]-labeled mannose from GDP-[^{14}C]Man to $AcPIM_1$ by scintillation counting and thin-layer chromatography. *B,* Saturation binding of GDP-Man or GDP to PimB' monitored by internal Trp fluorescence. *C,* Far-UV circular dichroism spectra of wild-type PimB' and of the H118S mutant.

Scheme 1 Schematic representation of steps of the S_Ni-like (internal return) reaction mechanism leading to retention of the α-configuration of the anomeric carbon of the donor saccharide. Adapted from reference (43).

TABLES

Table 1: Data acquisition and structure refinement statistics

Data acquisition	PimB' + GDP-Man	SeMet 1	SeMet 2	SeMet 3	SeMet 4	PimB' + Br-GDP-Man	PimB' + GDP-Man	PimB' + GDP
Beamline	ESRF ID29	Diamond I04	Diamond I04	ESRF ID23.1	ESRF ID29	Diamond I02	Diamond I04	ESRF ID14.2
Wavelength (Å)	0.9762	0.9805	0.9805	0.97905	0.97888	0.9201	0.9763	0.933
Space group	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å); a, b, c =	44.1, 50.0, 85.3	44.1, 50.1, 85.3	44.1, 50.0, 85.3	43.9, 49.9, 85.4	44.0, 50.0, 85.7	43.9, 49.9, 85.7	44.8, 87.0, 106.9	44.8, 86.3, 106.4
α, β, γ =	92.1, 92.7, 89.9	92.2, 92.6, 90.1	92.0, 92.8, 90.0	91.9, 92.4, 90.1	92.2, 92.6, 90.0	92.1, 92.7, 89.9	90, 90, 90	90, 90, 90
Resolution (Å)	44.0-2.2	29.96-2.90	29.91-2.90	25.38-2.90	29.97-2.80	29.91-2.90	37.29-2.0	45.37-2.23
High resolution shell (Å)	2.32-2.20	3.06-2.90	3.06-2.90	3.06-2.90	2.95-2.80	3.06-2.90	2.11-2.00	2.35-2.23
Rmerge (%)[1),]	7.3 (23.8)	4.9 (11.1)	7.1 (15.5)	6.7 (11.3)	8.5 (18.1)	11.6 (26.2)	9.6 (38)	9.3 (38.2)
Total number of observations[1)]	103582 (12352)	40261 (5462)	46289 (6804)	44487 (6484)	52162 (7446)	46058 (6569)	208274 (29919)	64840 (4871)
Unique reflections[1)]	35191 (4400)	13101 (1951)	13931 (2092)	13532 (2027)	17456 (2507)	15023 (2197)	28997 (4147)	19982 (2147)
$\langle I/\sigma(I) \rangle$[1)]	11.6 (5.9)	18.1 (7.1)	11.6 (5.3)	12.5 (7.8)	9 (5.2)	7.2 (3.2)	18.3 (5.7)	9.9 (2.3)
Completeness (%)[1)]	95.1 (81.2)	81 (83.1)	86.4 (89.3)	84 (86.4)	97.4 (95.8)	93.4 (93.4)	100 (99.8)	95.8 (72.5)
Multiplicity[1)]	2.9 (2.8)	3.1 (2.8)	3.3 (3.3)	3.3 (3.2)	3 (3)	3.1 (3.0)	7.2 (7.2)	3.2 (2.3)
Anomalous completeness (%)[1)]		70.5 (70.0)	84.5 (87.1)	77.3 (80.3)	92.5 (90.1)	82.4 (82.7)		
Anomalous multiplicity[1)]		1.7 (1.6)	1.7 (1.7)	1.7 (1.7)	1.5 (1.5)	1.7 (1.6)		
Δ_{ano} correlation between half-sets[1)]		0.483 (0.174)	0.272 (0.138)	0.265 (0.132)	0.246 (0.198)	0.101 (0.073)		

Table 1 (cont'd)

Refinement	PimB' + GDP-Man	PimB' + GDP-Man (orth)	PimB' + GDP (orth)
Space group	*P*1	$P2_12_12_1$	$P2_12_12_1$
Resolution range (Å)	44.0-2.2	37.3-2.0	45.4-2.4
Unique reflections	33243	27519	15991
R_{cryst} (%)	18.9	20.3	19
R_{free} (%)[2)]	22.1	23.2	23.6
No. of non-H atoms	6235	3068	3019
Protein	5829	2829	2851
Ligand	56	56	28
Solvent	350	183	140
RMSD bonds (Å)[3)]	0.008	0.007	0.007
RMSD angles (°)[3)]	1.15	1.02	1.09
Wilson B-factor ($Å^2$)	17.3	15.9	28
Aver. B-factor ($Å^2$)	13.1	14.2	20.4
Protein	12.2	13.8	20.1
Ligand	31.9	20.6	38.2
Solvent	17.5	19.3	24.7
RMSD B-factor between			
Bonded main chain atoms	0.42	0.26	0.34
Bonded side chain atoms	1.17	0.65	0.55
Ramachandran Plot[3)]			
Favoured (%)	98.7	98.1	98.0
Outliers (no.)	0	0	0

Table Footnotes: 1) high resolution shell in parentheses; 2) R_{free} was calculated based on 5% of reflections set aside for cross-validation; 3) deviations relative to ideal bond distances and angles in (53); 4) calculated using MOLPROBITY (54).

Table 2: Activity and saturation binding for PimB' and selected point mutants.

	GDP-Man K_d (μM)	**GDP K_d** (μM)	**Activity** (CPM)	**Residual Activity** (%)
WT	19.0 ±4.6	36.5 ±6.9	5528 ±979	100.0
Blank	-	-	30 ±22	0.5
N12A	20.4 ±6.2	21.3 ±5.5	1723 ±77	31.2
D13N	30.1 ±6.9	18.8 ±5.1	83 ±5	1.5
D13Y	-	-	16 ±0	0.3
D13A	42.8 ±11.9	61.3 ±17.4	104 ±5	1.9
G20W	21.9 ±6.4	46.5 ±40.2	176 ±15	3.2
I21S	13.4 ±7.0	26.5 ±7.4	134 ±22	2.4
I21A	56.0 ±14.5	40.9 ±8.6	110 ±5	2.0
Q22A	60.3 ±22.8	30.6 ±6.4	788 ±218	14.3
H118S	72.4 ±27.5	19.6 ±9.8	36 ±3	0.7
H120S	-	-	4490 ±524	81.2
G123P	29.3 ±5.9	42.3 ±9.3	4784 ±298	86.6
S205G	18.7 ±4.6	33.6 ±6.9	6609 ±1207	119.6
R206S	42.7 ±9.4	72.4 ±17.2	311 ±47	5.6
R210S	40.2 ±10.7	44.1 ±28.8	481 ±58	8.7
K211Q	30.5 ±17.6	40.3 ±14.6	193 ±25	3.5
E290Q	35.7 ±18.9	41.7 ±21.5	252 ±30	4.6
E290D	55.0 ±17.7	322.7 ±518.7	162 ±14	2.9
E290N	NB	NB	59 ±16	1.1
G291S	68.7 ±35.5	74.2 ±37.2	292 ±241	5.3

Hyphen (-) = binding not assayed. NB = no binding detected.

Figure 1

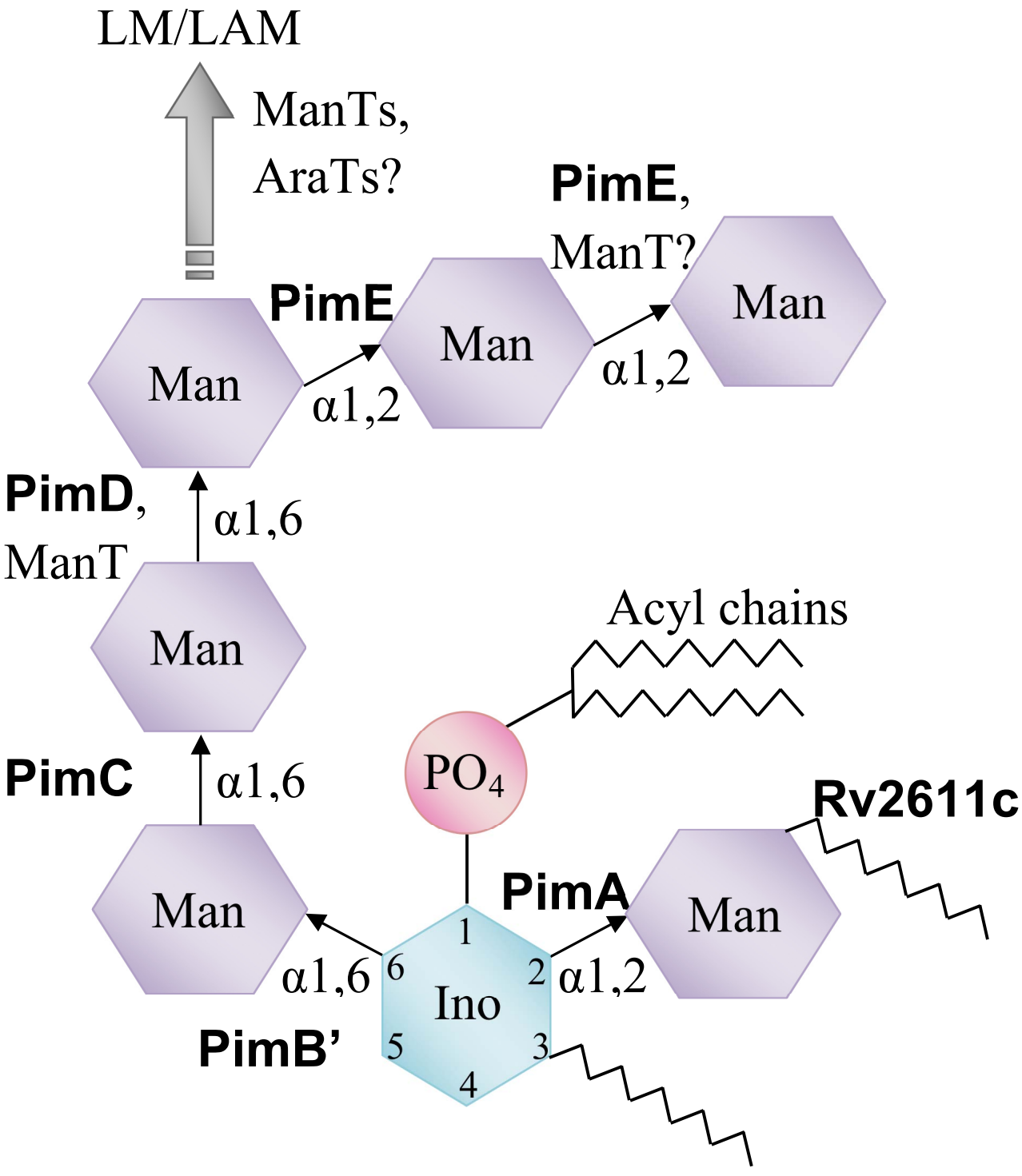


Figure 2

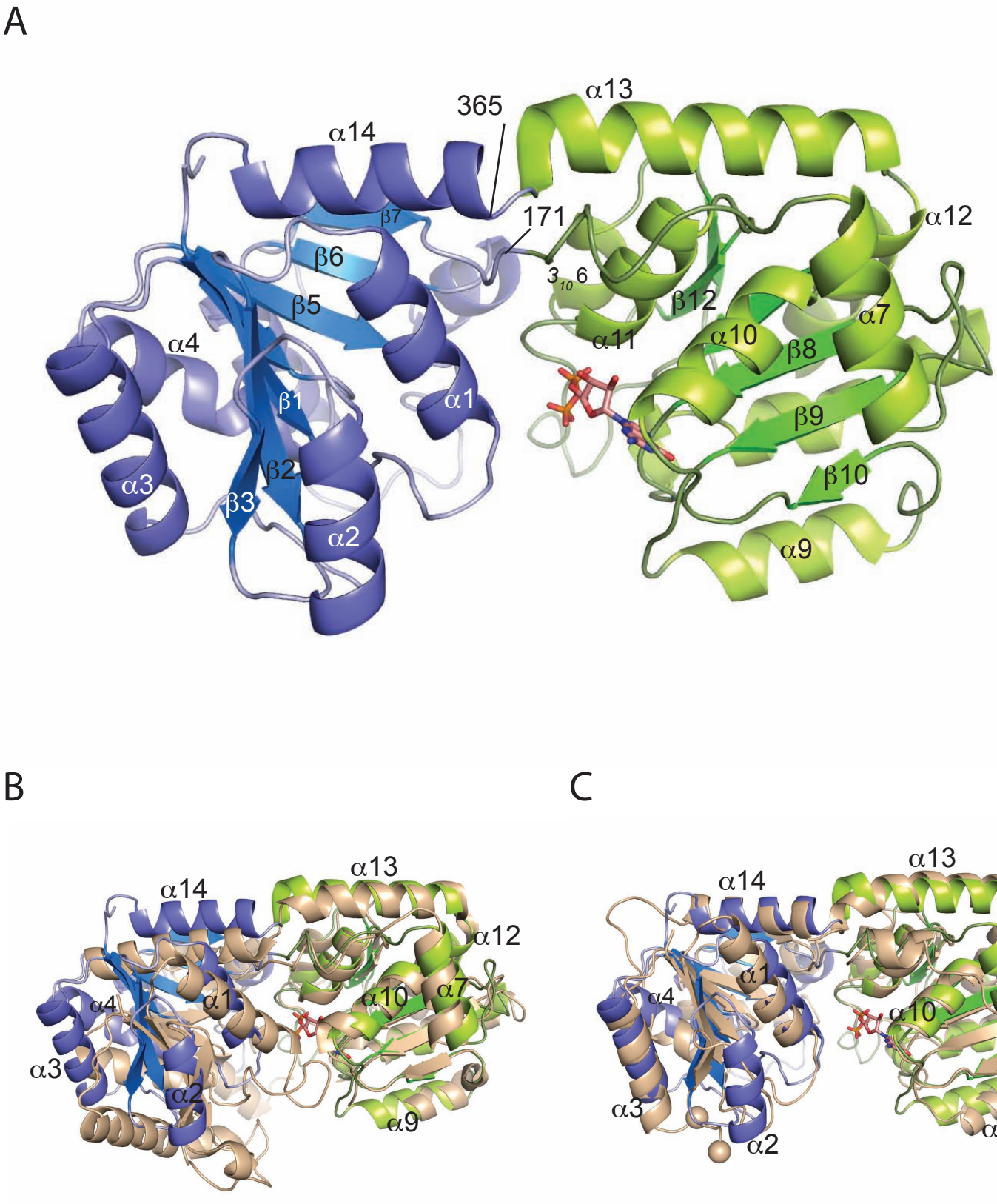


Figure 3

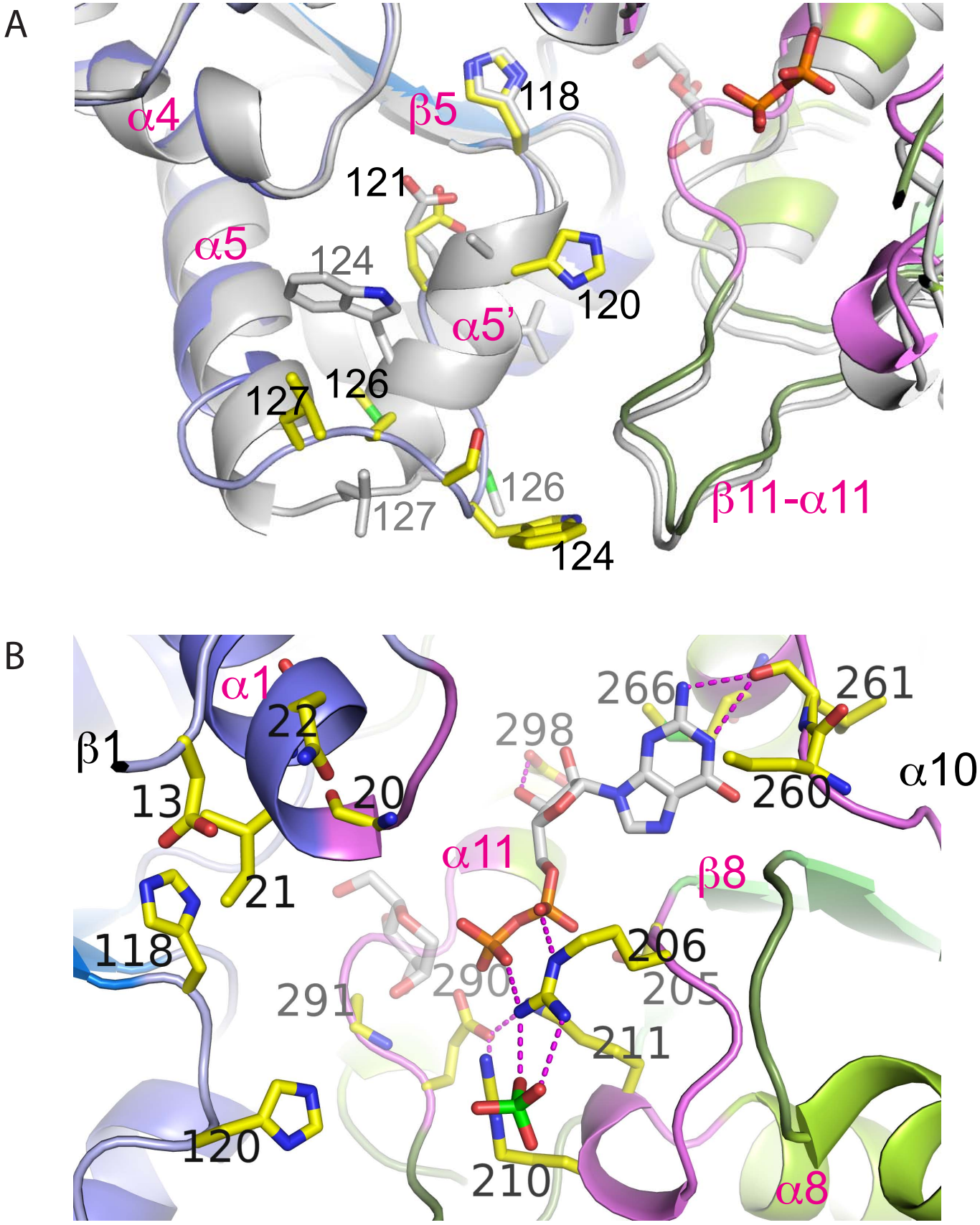


Figure 4

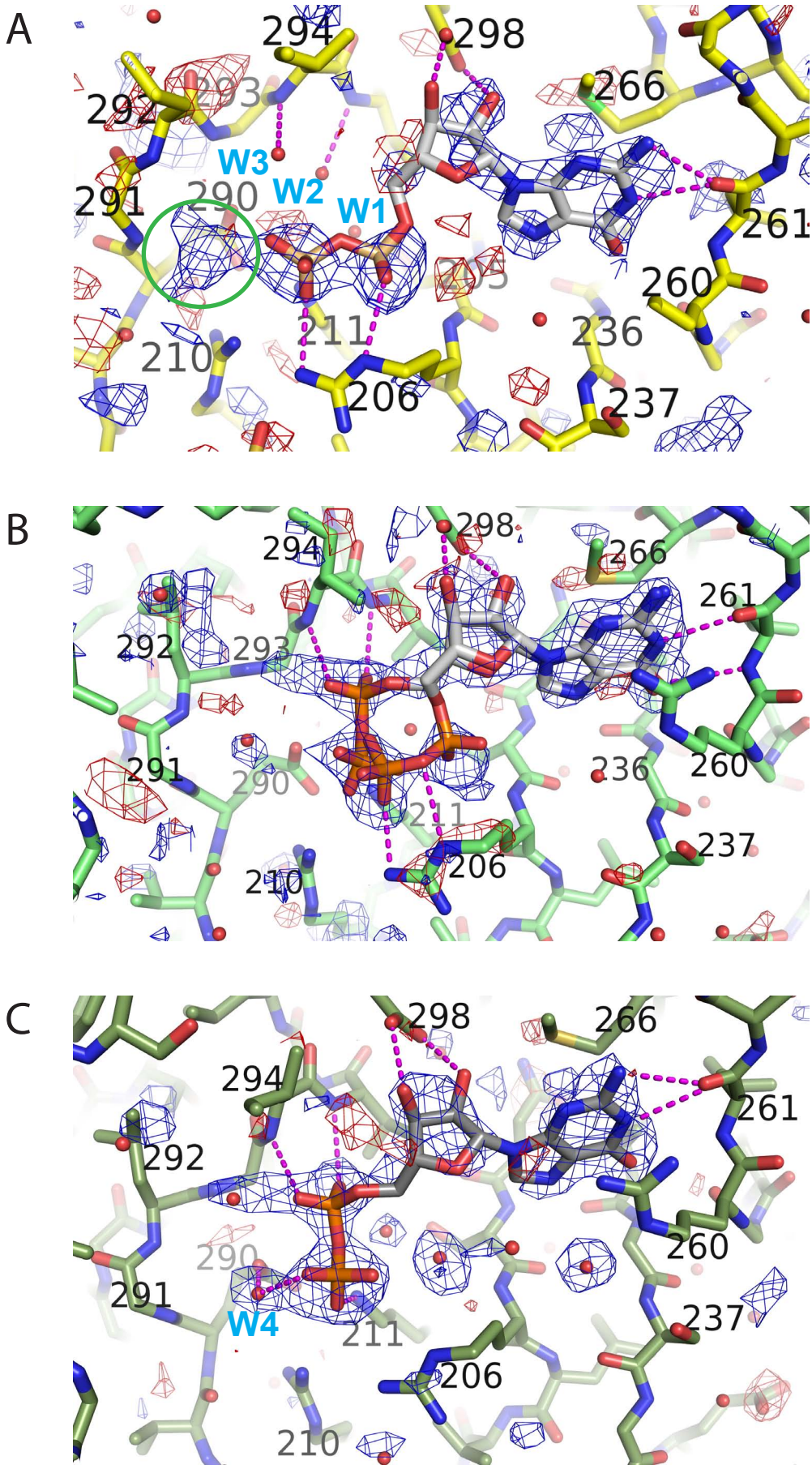
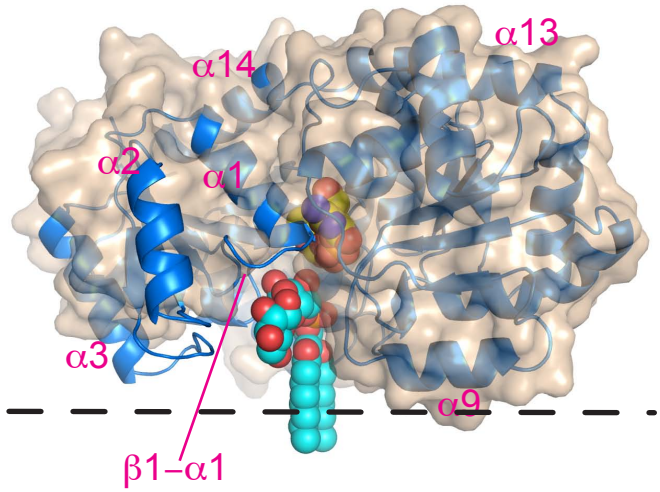


Figure 5

A



B

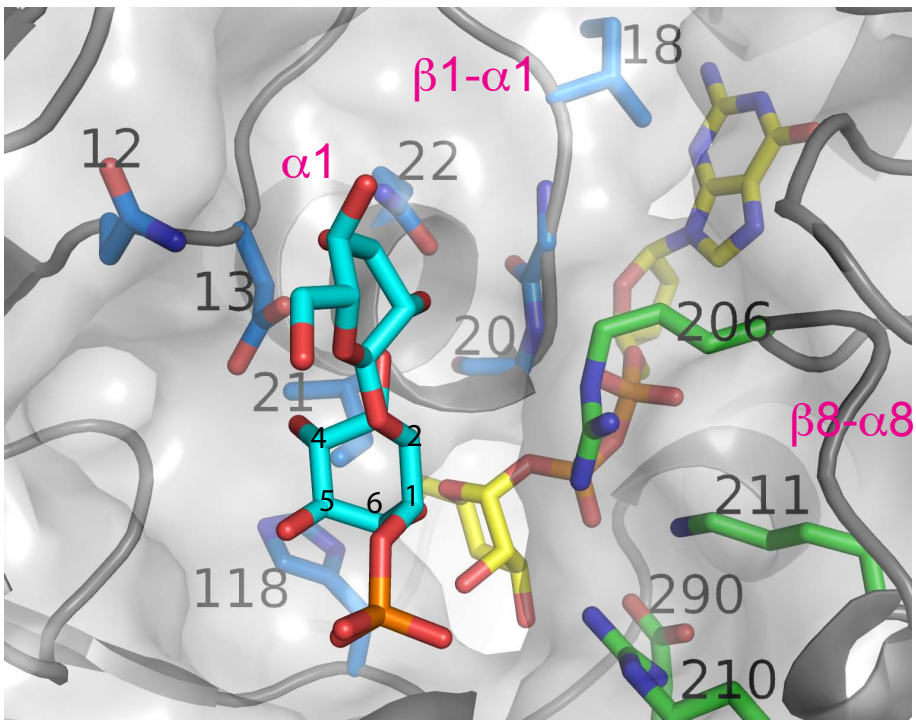
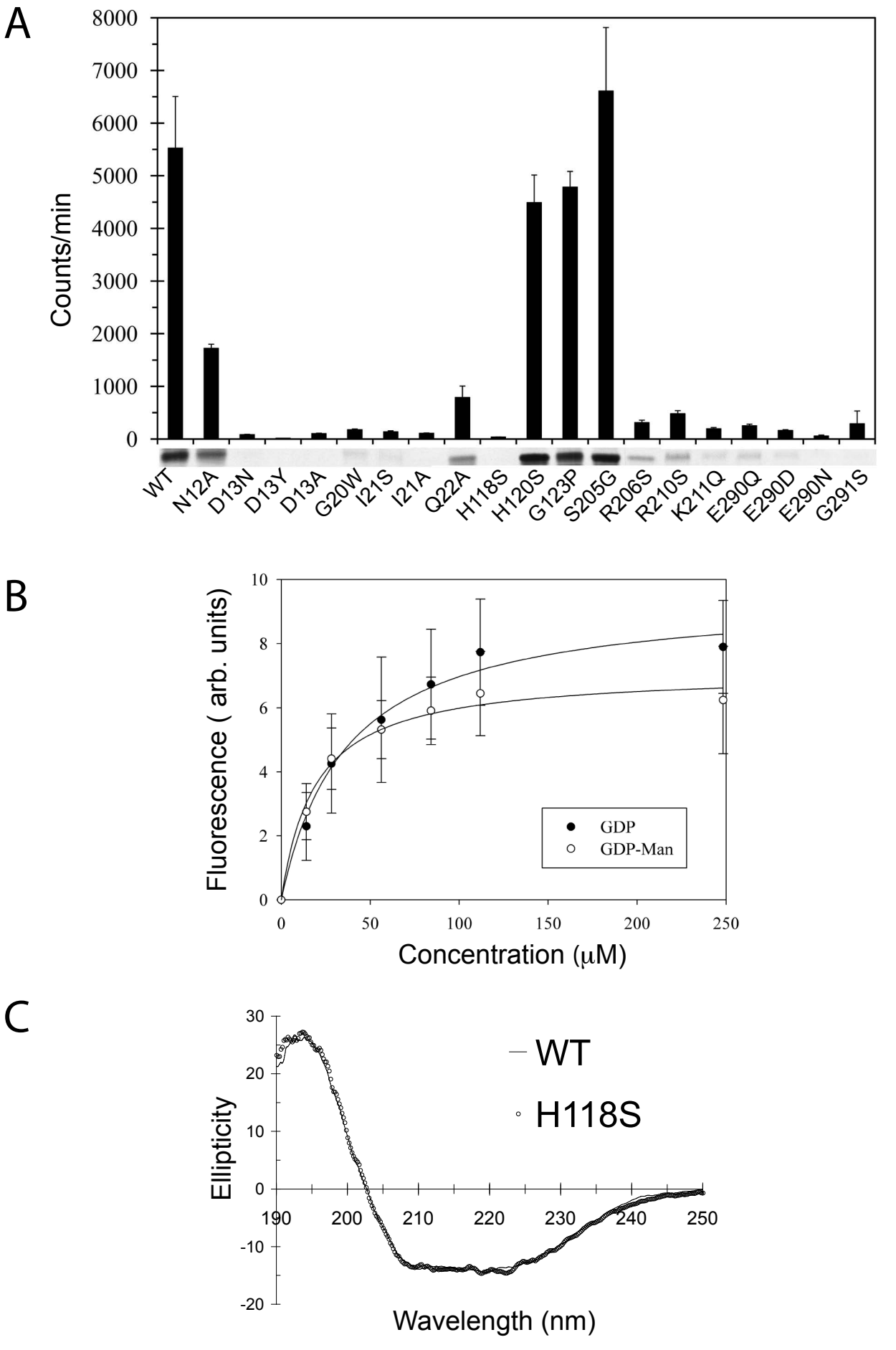


Figure 6



Scheme 1